WHAT IS CLAIMED IS:

1   1.   A method for handling redirects in documents, comprising:

2   forming at least one equivalence class that includes documents that are connected

3 through a redirect;

4   detecting cycles for each equivalence class, wherein documents in a cycle are

5 marked so that they are not indexed;

6   detecting incomplete chains for each equivalence class, wherein documents in an

7 incomplete chain are marked so that they are not indexed; and

8   selecting a representative for each equivalence class.

1   2.   The method of claim 1, wherein the representative is selected based on a

2 type of redirect in an equivalence class.

1   3.   The method of claim 1, wherein the representative is selected based on a

2 rank of each document in the equivalence class.

1   4.   The method of claim 1, further comprising:

2   locating each document that contains a redirect; and

3   creating an entry in a redirect file for each document.

1   5.   The method of claim 4, wherein the entry includes a source path, a target

2 path, and a redirect type.

1   6.   The method of claim 1, further comprising:

2   detecting duplicate documents in two different equivalence classes; and

3   merging the equivalence classes.

1   7.   The method of claim 6, wherein documents are duplicates if a certain

2 portion of their content is similar.

1   8.   The method of claim 1, wherein the documents in the at least one

2 equivalence class include a target document and one or more source documents and

3 wherein the selected representative is one of the source documents, further comprising:

4   propagating the content of the target document to the selected representative.

1   9.   The method of claim 1, wherein the documents in the at least one

2 equivalence class include a target document and one or more source documents, and

3 wherein at least one source document includes a path to the target document.

1   10.   The method of claim 9, further comprising:

2   indexing the content of the target document with a path of the representative.

1   11.   The method of claim 1, wherein marking documents so that they are not

2 indexed includes marking documents to indicate the documents are to be ignored.

1   12.   The method of claim 1, further comprising:

2   determining a rank for each of the documents, wherein the rank represents an

3 importance of each document relative to the other documents.

1   13.   An article of manufacture including a program for handling redirects in

2 documents, wherein the program causes operations to be performed, the operations

3 comprising:

4   forming at least one equivalence class that includes documents that are connected

5 through a redirect;

6   detecting cycles for each equivalence class, wherein documents in a cycle are

7 marked so that they are not indexed;

8   detecting incomplete chains for each equivalence class, wherein documents in an

9 incomplete chain are marked so that they are not indexed; and

10   selecting a representative for each equivalence class.

1   14.   The article of manufacture of claim 13, wherein the representative is

2   selected based on a type of redirect in an equivalence class.

1   15.   The article of manufacture of claim 13, wherein the representative is

2   selected based on a rank of each document in the equivalence class.

1   16.   The article of manufacture of claim 13, wherein the operations further

2   comprise:

3   locating each document that contains a redirect; and

4   creating an entry in a redirect file for each document.

1   17.   The article of manufacture of claim 16, wherein the entry includes a

2   source path, a target path, and a redirect type.

1   18.   The article of manufacture of claim 13, wherein the operations further

2   comprise:

3   detecting duplicate documents in two different equivalence classes; and

4   merging the equivalence classes.

1   19.   The article of manufacture of claim 18, wherein documents are duplicates

2   if a certain portion of their content is similar.

1   20.   The article of manufacture of claim 13, wherein the documents in the at

2   least one equivalence class include a target document and one or more source documents

3   and wherein the selected representative is one of the source documents, wherein the

4   operations further comprise:

5   propagating the content of the target document to the selected representative.

1   21.   The article of manufacture of claim 13, wherein the documents in the at

2   least one equivalence class include a target document and one or more source documents,

3   and wherein at least one source document includes a path to the target document.

1   22.   The article of manufacture of claim 21, wherein the operations further

2   comprise:

3   indexing the content of the target document with a path of the representative.

1   23.   The article of manufacture of claim 13, wherein the operations for

2   marking documents so that they are not indexed include operations for marking

3   documents to indicate the documents are to be ignored.

1   24.   The article of manufacture of claim 13, wherein the operations further

2   comprise:

3   determining a rank for each of the documents, wherein the rank represents an

4   importance of each document relative to the other documents.

1   25.   A computer system including logic for handling redirects in documents,

2   comprising:

3   forming at least one equivalence class that includes documents that are connected

4   through a redirect;

5   detecting cycles for each equivalence class, wherein documents in a cycle are

6   marked so that they are not indexed;

7   detecting incomplete chains for each equivalence class, wherein documents in an

8   incomplete chain are marked so that they are not indexed; and

9   selecting a representative for each equivalence class.

1   26.   The computer system of claim 25, wherein the representative is selected

2   based on a type of redirect in an equivalence class.

1   27.   The computer system of claim 25, wherein the representative is selected
2 based on a rank of each document in the equivalence class.

1   28.   The computer system of claim 25, wherein the logic further comprises:
2   locating each document that contains a redirect; and
3   creating an entry in a redirect file for each document.

1   29.   The computer system of claim 28, wherein the entry includes a source
2 path, a target path, and a redirect type.

1   30.   The computer system of claim 25, wherein the logic further comprises:
2   detecting duplicate documents in two different equivalence classes; and
3   merging the equivalence classes.

1   31.   The computer system of claim 30, wherein documents are duplicates if a
2 certain portion of their content is similar.

1   32.   The computer system of claim 31, wherein the documents in the at least
2 one equivalence class include a target document and one or more source documents and
3 wherein the selected representative is one of the source documents, wherein the logic
4 further comprises:
5   propagating the content of the target document to the selected representative.

1   33.   The computer system of claim 25, wherein the documents in the at least
2 one equivalence class include a target document and one or more source documents, and
3 wherein at least one source document includes a path to the target document.

1   34.   The computer system of claim 33, wherein the logic further comprises:
2   indexing the content of the target document with a path of the representative.

1   35.   The computer system of claim 25, wherein marking documents so that

2 they are not indexed includes marking documents to indicate the documents are to be

3 ignored.

1   36.   The computer system of claim 25, wherein the logic further comprises:

2   determining a rank for each of the documents, wherein the rank represents an

3 importance of each document relative to the other documents.